Module 1

Vikas Jadhav Y, Ananya Verma

February 2025

Activity 1

1a

The data was generated in the following ways.

$$x \in \mathcal{U}(0, 1) \tag{1}$$

$$y = sin(2\pi x) + \mathcal{N}(\mu = 0, \sigma = 0.3) \tag{2}$$

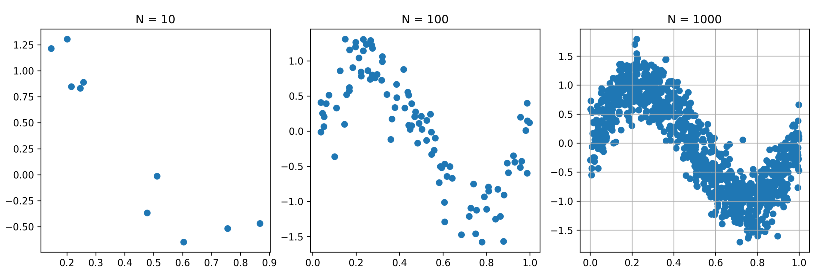

Figure 1: Figure showing the samples drawn with N = 10, 100, and 1000

1b

The data was fit with polynomials using linear regression. We can observe that for smaller degrees of the polynomial, no matter how many data points we have, it is always underfit. However, for higher degree polynomials, there is overfitting in the cases where the number of data points are low. There are good fits in the cases where there is both higher degree and large data points.

For the case where we fit the $N = 100$ data points with a 9th degree polynomial, we obtained an $E_{RMS} = 0.28$. This is very close to the noise RMS of $\sigma = 0.3$ showing that it is a very good fit since the only deviations come from the very nature of the added noise.

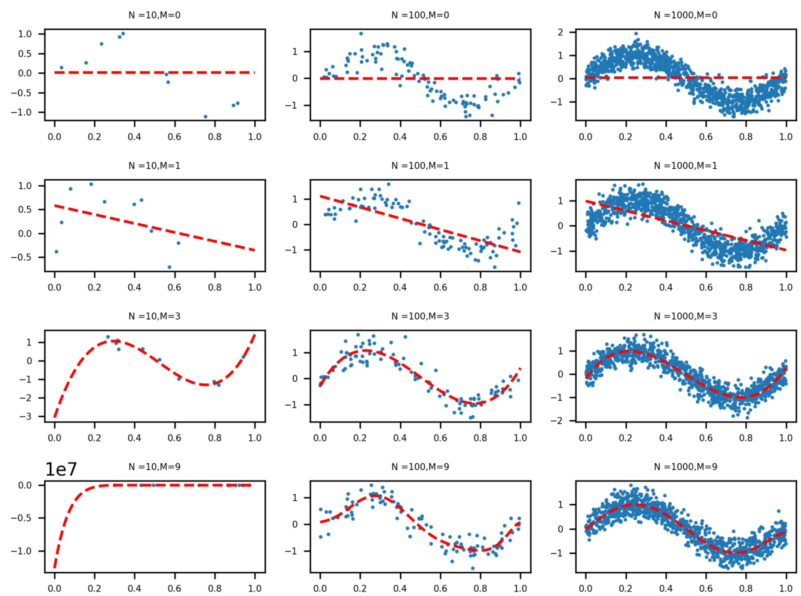

Figure 2: Fits of polynomial with degree M for various cases of N data points

Activity 2

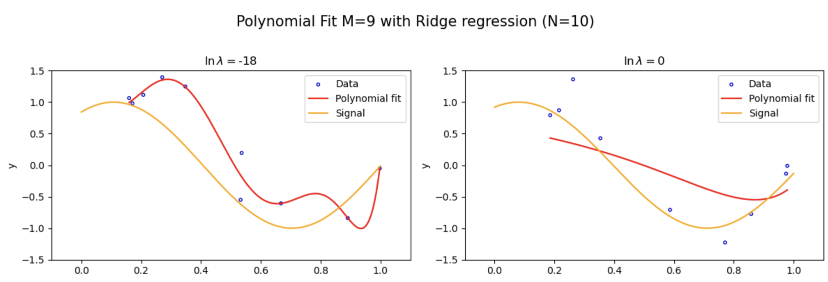

Figure 3:

2a

Linear regression with regularization helps to overcome the overfitting problem by introducing a penalty term λ in the error function. The new error function becomes:

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2 + \frac{\lambda}{2} * \|w\|^2 \quad (3)$$

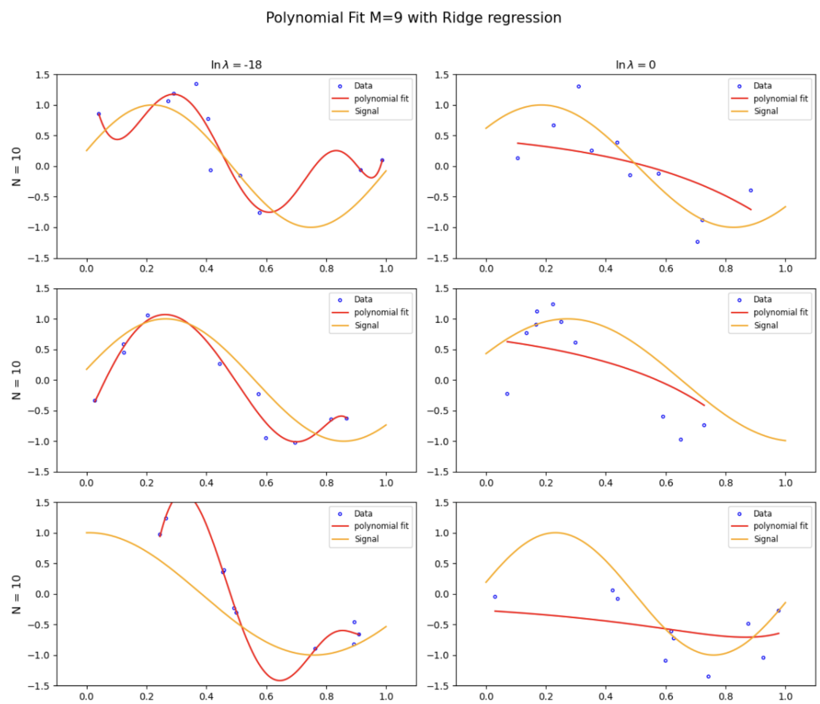

Figure 4:

By minimizing this error function we get the best fitted parameters

$$\sqsupseteq *$$

. If

$$\lambda = \prime$$

, then

$$\mathcal{E}(\sqsupseteq)$$

is our normal linear regression error function which shows overfitting problem for higher order polynomial with small sample of data points. But if we increase the

$$\lambda$$

values, the wild oscillation (fig 3a) decreases and givs us the following plots.

2b

Our E_{RMS} plot in Fig. 6 here shows the expected trend. The E_{RMS} curves for test sets decreases and increases eventually; the training set is always lower than the test set.

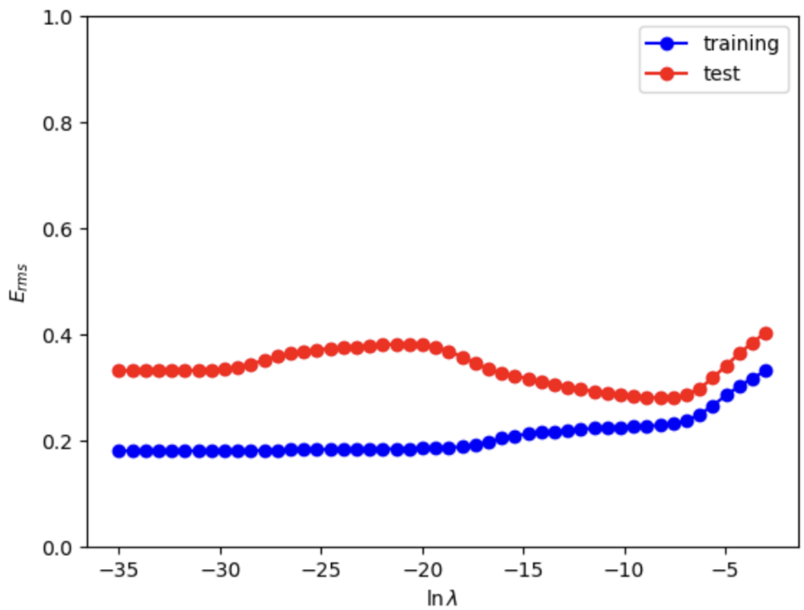

Figure 5:

Activity 3

The given probabilities are

$$P(A) = 0.001 \rightarrow P(\neg A) = 0.999 \tag{4}$$

$$P(+|A) = 0.98 \tag{5}$$

$$P(+|\neg A) = 0.03 \tag{6}$$

We are asked the probability of having the disease, given a positive test, i.e $P(A|+)$

According to Bayes theorem, this is given by,

$$P(A|+) = \frac{P(+|A)P(A)}{P(+)} \tag{7}$$

In this equation, we can make the following identifications

1. Likelihood - $P(+|A)$
2. Prior - $P(A)$
3. Normalisation - $P(+)$

The normalisation terms is obtained by summing over the cases of having the disease and not having the disease. Thus,

$$P(+) = P(+|A)P(A) + P(+|\neg A)P(\neg A) \tag{8}$$

Using this form of normalisation in equation(6) gives us $P(A|+) = 0.03$.

Activity 4

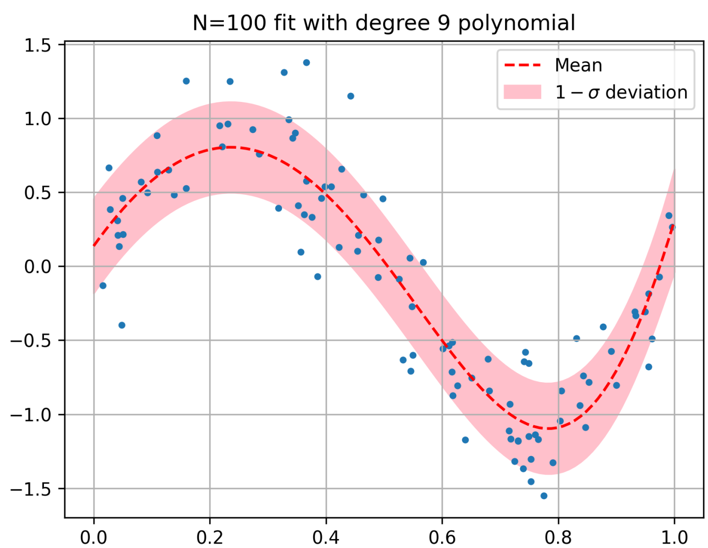

Figure 6: Bayesing fit for N = 100 with a 9th degree polynomial

4a

Here $\beta = 1/\sigma^2$ of the noise. Since $\sigma = 0.3$, we get a value of $\beta = 11.1$

4b

In the frequentist approach, we only get one predication for a given value of the independent parameter. However, in the bayesian approach, we get a distribution of predictions for a given value of the independent parameter. This approach is thus a better approach since it takes into account different noise realizations from the same noise rms into account. In the frequentist approach, a different noise realisation would lead to a completely new fit, howerver, the bayesian fit is more robust.

Our understading is that frequentist fits for different realisations of the same noise, when overlaid on each other, will replicate the $1-\sigma$ band that the bayesian fit produces.